

DockingShop: A Tool for Interactive Molecular Docking

Ting-Cheng Lu¹, Nelson L. Max^{2, 3}, Jinhui Ding¹, E. Wes Bethel³, Silvia N. Crivelli^{1, 3}

ABSTRACT

Given two independently determined molecular structures, the molecular docking problem predicts the bound association, or best fit between them, while allowing for conformational changes of the individual molecules during construction of a molecular complex. DockingShop is an integrated environment that permits interactive molecular docking by navigating a ligand or protein to an estimated binding site of a receptor with real-time graphical feedback of scoring factors as visual guides. Our program can be used to create initial configurations for a protein docking prediction process. Its output --the structure of a protein-ligand or protein-protein complex-- may serve as an input for a protein docking algorithm, or an optimization process. This tool provides molecular graphics interfaces for structure modeling, interactive manipulation, navigation, optimization, and dynamic visualization to aid users steer the prediction process using their biological knowledge.

CR Categories: J.3 [Life and Medical Sciences]: Biology and Genetics; I.3.6 [Computer Graphics]: Methodology and Techniques-Interaction Techniques; I.3.8 [Computer Graphics]: Applications; I.6.8 [Simulation and Modeling]: Types of Simulation-Visual; I.6.9 [Simulation and Modeling]: Visualization-Information Visualization

Keywords: Bioinformatics Visualization, Protein-Protein Interaction, Molecular Visualization, Interactive Visualization, Computational Science, Interaction, Applications of Volume Graphics and Volume Visualization

1 INTRODUCTION

One of the grand challenges in computational biology is the molecular docking problem. The molecular docking problem is to determine how molecules interact with other molecules and plays a key role in understanding how cells function. Its solution will help scientists find better ways of changing cell functions with new and more specialized drugs designed specifically for that purpose [See Figure 1]. Experimental efforts have been devoted to the analysis of intermolecular interactions that form many putative complexes. However, the experimental determination of protein-protein complex structures remains difficult. Therefore, a

number of computational approaches have been proposed. In general, solving the "docking" problem can be treated as a problem of finding the low-energy binding modes of a ligand within the active site of a receptor whose structure is known. Thus, computational methods for molecular docking require an accurate representation of the molecular energetics as well as an efficient minimization algorithm to search the potential binding modes. Most docking algorithms have been developed for two main classes of crystal structures of complexes deposited in the Protein Data Bank: enzyme-inhibitor and antibody-antigen complexes. However, there are other intermolecular interactions that involve molecules of different types and sizes that are desirable to study and predict. Furthermore, a thorough assessment of current methods for protein docking [5] refers to the need for better scoring functions and more effective mechanisms for handling conformational flexibility. Both problems are related to the physical chemistry of proteins and they require a realistic representation of the protein structure and of all the free energy terms that are involved in folding and assembly [16].

The docking problem can be classified as bound or unbound based on whether a protein complex is extracted from a compound of molecules or it is made of individually solved structures. Unbound docking is far more complex because it requires computational schemes to reconstruct a complex using unbound structures. This additional complexity derives from conformational changes between the bound and unbound structures that take place during the binding process [17]. The challenge is the conformational change upon binding which involves significant backbone movement [12].

Most docking algorithms have adopted a two-stage approach. In the first stage, called *conformational space search*, both molecules are treated as rigid bodies and their relative position and orientation are fully explored with six degrees of freedom. The rigid body transformation of ligand and receptor based on Fourier transformation has been widely applied in the first stage of the docking simulation. However, such an approach does not allow users to inspect and assess possible solutions in real time and is computationally intensive. In fact, it might take days to find solutions for this stage due to the exhaustive conformational space search procedure inherent to the FFT process. Moreover, this approach is less efficient when the sampling space is large like in cases in which the whole surface of a receptor is sampled and all possible combinations of patches of molecular surface between receptor and ligand need to be generated. The computational problem is even more profound when considering the protein flexibility and the increasing demand to screen large databases of proteins and/or potential drugs [17].

In the second stage, called *ranking of potential solutions*, tens to thousands of potential conformational structures obtained in the first stage are refined and re-ranked by using more detailed scoring functions to discriminate between native and non-native docked conformations and to find the lowest-energy binding mode of the ligand to the receptor. The scoring functions use various criteria, including geometric and chemical complementary measures, hydrogen bonds, and free energy, independently or in combination. The use of biochemical and biological information

¹California Institute for Quantitative Biomedical Research, Berkeley, California, tlu@lbl.gov, jhding@berkeley.edu

²Department of Computer Science, University of California, Davis, max@cs.ucdavis.edu

³Lawrence Berkeley National Laboratory, Berkeley, California, {snrivelli, ewbethel}@lbl.gov

plays a key role in identifying the native binding modes for the correctly predicted targets. However, although some algorithms can rank correct solutions for some predictive docking cases, for most complexes the highest ranked structures are still false positives [17]. Therefore, scientists must visually inspect the predicted solutions that result from the second stage of a docking algorithm, usually consisting of many large coordinate files, using one of the many existing commercial or academic molecular graphics packages. Our goal is to develop a tool that provides support to different docking methods rather than a docking method itself. DockingShop provides an interactive molecular docking environment that includes real-time visual feedback to aid in steering the docking process for rapid estimation of the conformational binding mode along with flexibility of side chain and backbone movement. Its features include 1) sculpting of molecular structures, 2) determining molecular interactions using visual feedbacks to steer the docking process in real time and 3) modeling and visualizing the details of specific interactions at the atomic level. Our tool integrates human intuition and biological knowledge to steer the prediction process. Our work also advances state-of-the-art computational biology and scientific visualization by directly integrating several existing forms of visualization.

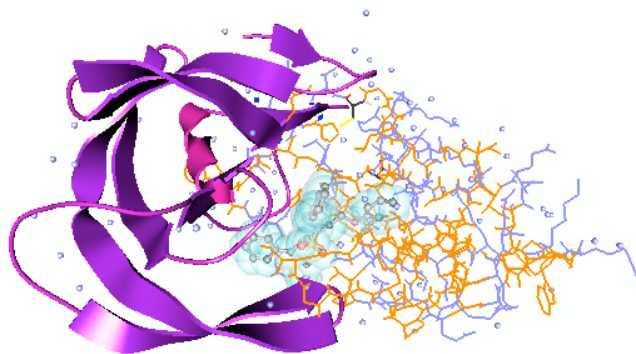


Figure 1. DockingShop rendering of HIV-1 protease in complex with the cyclic sulfamide inhibitor Aha006 (PDB code 1AJV).

2 RELATED WORK

Most molecular graphics tools for molecular docking are rendering tools. Examples of those packages are VMD [1], Chimera [21], and PyMol [20]. Chimera and PyMol allow users to perform some limited molecular manipulations but they do not support interactive mechanisms for docking. These packages are used for post processing and are not tightly integrated into the docking process. A few packages provide some interactive features: HotDock [9] permits users to handle a ligand like a space shuttle and navigate it towards its "landing position" at the prospective binding site. An efficient collision detection algorithm calculates bounding spheres based on derived secondary structure information for the involved molecules. It generates acoustical feedback and the colliding atoms are highlighted when they "crash". The purpose of HotDock is to provide a navigation system to dock a ligand. However, many useful properties such as geometric complementarities, hydrophobic effect, hydrogen bonds, and energy function for determination of the conformational binding mode are missing. Insight II from Accelrys [10] is a molecular modeling environment that helps to find docking structures by manually moving the ligand to a position where the energy is at its lowest, while the interaction energy is calculated based on an energy grid. Accelrys also

provides a more thorough and rigorous product called Affinity, which was designed for protein-ligand docking only. Insight is ideally suited for a rapid estimation of the ligand binding to the receptor, but it is difficult to determine exactly how the ligand and receptor interact. Therefore, incorporation of experimental data and consideration of conformational changes during the binding process is performed using Affinity. Although Affinity's use of a full force field energy expression allows for accurate representation of intermolecular interaction, there is no flexibility of backbone. Another interesting program is IMD [3], a novel simulation method for interactive manipulation of bio-molecular systems to probe a molecular dynamic process. It combines the molecular visualization program VMD with a molecular dynamic program NAMD [2] and supports manipulation of molecules by applying force to single atoms with real-time force feedback. IMD combines interactive molecular visualization with a background MD simulation program thus providing flexibility for users to monitor the molecular dynamic process through a visual display. However, IMD is a system focused on molecular dynamic simulation. In addition, IMD and Insight do not permit to use and update biochemical properties to realize the interaction between molecules in real-time. Also, these programs do not support the integration of different biochemical properties for an interactive scoring process. Furthermore, none of these systems provides enough flexibility of side chains and backbone to perform conformational changes of molecular structures; this limitation reduces the chances for success and accuracy of the predictions in many docking algorithms [12]. Some docking programs [8,23] integrate tactile sense technology in the docking process but their extra computational cost makes them difficult to use. Another weakness with haptic interfaces is that the haptic devices give only point-samples, which is akin to searching through a dark room with a laser beam rather than a floodlight. Furthermore, they do not support the integration of scoring functions or the flexibility of side-chains and backbone. ICM-Pro from MolSoft provides a GUI interface that combines a step-by-step docking menu with other modules like ICM-Docking to perform docking simulation with flexibility of side-chains and backbone. It uses Internal Coordinate Mechanics [11] to describe intra- or inter-molecular geometry. However, ICM-Docking performs docking using the ICM stochastic global optimization procedure and lacks mechanisms for integrating other procedures or for allowing user interaction. The scoring function is not coupled in the docking process and final visual inspections are done on ICM-Pro, a molecular graphic program. VRDD [22] enables molecular visualization and interactive docking without flexibility of backbone in a VR environment. VRDD supports side chain conformational search by allowing the users to select any side chain dihedral angles. Binding free energies are computed in real time and used in numerical format to guide the docking process. Although VRDD uses the solvent accessible surface representation to decrease visual complexity, users still need to identify the binding sites based on general shape and surface coloring without other visual guides to determine potential active sites (pockets). This method of determining binding sites decreases the accuracy and only experienced users with strong biology knowledge might be able to achieve this task. The level of atomic overlap is indicated by the change in the residue's color. However, other important factors commonly used for scoring functions like hydrogen bonds and hydrophobic effects are neither taken into account nor visualized.

In this paper, we present DockingShop, an integrated graphical environment that is specifically designed to provide support for molecular docking by offering the following capabilities: interactive navigation of molecules, adjustable scoring functions, visualization of elements in the scoring scheme, integration of

user-selected energy and scoring functions, flexibility of molecular structures and steering mechanisms based on user knowledge and intuition. We will describe these features in the next section. The development of our program is driven by the concept of computer aided molecular design (CAMD) to enable users to simulate the structure, behavior, and interactions of molecules on computers. DockingShop is the successor of ProteinShop [7], which was initially developed to help users predict the 3D dimensional structure of new folds. ProteinShop creates a variety of partially or fully folded initial configurations to be used as seeds for further minimization of a physics-based energy function [4]. Furthermore, because the energy functions are far from perfect, ProteinShop can incorporate user-provided scoring functions to filter the better configurations from the bad ones. In addition, DockingShop provides an interactive docking environment for a rapid estimation of the conformational binding mode with flexibility of side chains and backbone movement. It permits the visualization of parameters used in scoring functions as indexes to evaluate the quality of the models and as guides to steer the docking process. The manipulation of molecular structures provides further exploration of conformational space. The integration of biological knowledge and user intuition provides a useful environment to support interactive real-time simulation of molecular interactions at the atomic level.

3 IMPLEMENTATION

The design of DockingShop follows a modular development concept. DockingShop is written in C++ and relies on the OpenGL library for three dimensional graphics rendering and is portable across most Unix platforms like Linux, SGI, and Mac. DockingShop uses FLTK, an interpreted interface layer, for the graphical user interface. Like ProteinShop, DockingShop provides support for modeling and visualization of molecules and biological data, constraints-based manipulation of molecular structure based on inverse kinematics algorithms, and optimization of an energy function. However, the essential elements of computational biology include computational methods to predict structure, properties, and behavior of those molecules along with molecular modeling and visualization. To that end, we are developing interface layers to bridge between computational biologists and emerging computational methods. Such interfaces allow developers to plug in their own codes to build powerful new applications for computational biology research.

During interactive docking, DockingShop provides six degrees of freedom for molecular manipulation. Users can freely move the whole complex of molecules or an individual molecule inside the complex using either the pilot window or the six-dial pop-up panel provided by the GUI. The user can also manipulate the molecular structure by dragging with the mouse a 3D manipulation widget, a translucent green box surrounding the secondary structure selected by a user. The molecular structure changes shape when the widget is translated or rotated. (See Section 4.4 for details.) Other steering mechanisms for molecular docking are described in the next sections.

4 SIMULATION OF MOLECULAR INTERACTIONS

DockingShop supports exploratory and interactive steering of molecular docking by leveraging user knowledge through integrating interactive visualization with real-time simulations. Next, we provide a detailed description of how our program supports molecular docking prediction. We describe scoring function visualization in Section 4.1 and scoring functions integration in Section 4.2. We then illustrate the interactive docking steps and show how scoring functions can guide the docking process in Section 4.3. Finally, we discuss flexibility of

molecular structures in Section 4.4 and energy minimization for refinement of structures in Section 4.5.

4.1 Visualization of Molecular Interaction

The search algorithms used in the conformational space search stage may produce a very large number of possible solutions that are unmanageable for any practical purposes. Therefore, a reliable set of scoring functions is required to discriminate between good candidates --those having low RMSD (root mean square deviation) from the native complex-- and bad ones within a reasonable computational time. The choice of parameters used for scoring functions depends on the breadth of sampling: entire surface vs. potential binding sites. These important computational parameters include hydrogen bonds, free energy, geometric complementarities, inter- and intra-molecular overlap, and hydrophobic effect [17]. Our program visualizes these parameters as "live" guides to help users navigate a ligand or protein to the potential conformational binding mode, to understand the behavior of a protein structure during the molecular interaction, and to discriminate between native and non-native conformations.

An extremely important feature of DockingShop is that it permits users to easily couple their own scoring functions so that they may search for a solution guided by scoring functions of their choice. This allows for an extremely level of flexibility and underscores one of the main capabilities of our program: it allows its users to incorporate their knowledge into the simulation process. In this section we describe visualization of the scoring function parameters supported by DockingShop. In Section 4.2 we describe how the users may combine these parameters to design scoring functions.

4.1.1 Hydrogen Bonds Visualization

Hydrogen bonds play an important role in molecular docking because they help to stabilize and strengthen a bound complex. Many molecular docking methods use the number of hydrogen bonds as a measurement to distinguish between correct structures and false solutions. A hydrogen bond is an interaction between a covalently bonded hydrogen atom on a donor group (i.e., O-H or N-H and a pair of non-bonded electrons on an acceptor group (i.e., :O=C or :N≡) as shown in Figure 2. The hydrogen bond length is the distance from the center of the donor atom to the center of the acceptor atom.

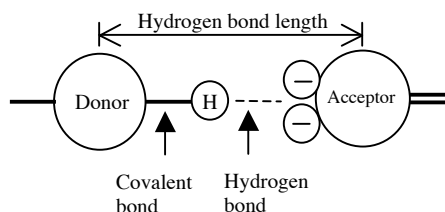


Figure 2. Hydrogen bond between Donor and Acceptor.

DockingShop visualizes inter- and intra-molecular hydrogen bonds. Moreover, it provides visual guides to facilitate the formation of hydrogen bonds through molecular manipulations. The program renders a *bond site*, i.e., the midpoint of a hypothetical hydrogen bond, for each charged backbone group, showing a potential bond's midpoint position and orientation. Forming bonds is thus reduced to aligning the midpoints and orientations of two differently charged backbone groups. The program constantly monitors the position and orientation of hydrogen bonding sites along the backbone, and renders a dashed yellow line as an intra-molecular guide between all pairs of negatively charged C=O and positively charged N-H groups that satisfy certain constraints. Likewise, the program also constantly

monitors the position and orientation of hydrogen bonding sites between molecules and renders a dashed green line as an inter-molecular guide between all pairs of donor and acceptor groups that satisfy certain constraints. Both constraints for inter- and intra-molecular hydrogen bonds are based on a simplified geometrical model to determine whether two opposing charged groups form a hydrogen bond. The two groups form a hydrogen bond if they are within certain distance and the direction of the bond satisfies and angle constraint. In addition, DockingShop renders hydrogen bonds that may form between the docking molecules and the water molecules that may be in the bonding areas. These hydrogen bonds are represented with a dashed blue line. A transparent green sphere highlights the area surrounding a potential hydrogen bond between molecules. These spheres help users identify the locations of hydrogen bonds among all rendered details. The length of hydrogen bond is also shown along the dashed line and continuously updated during the interaction (Figure 3).

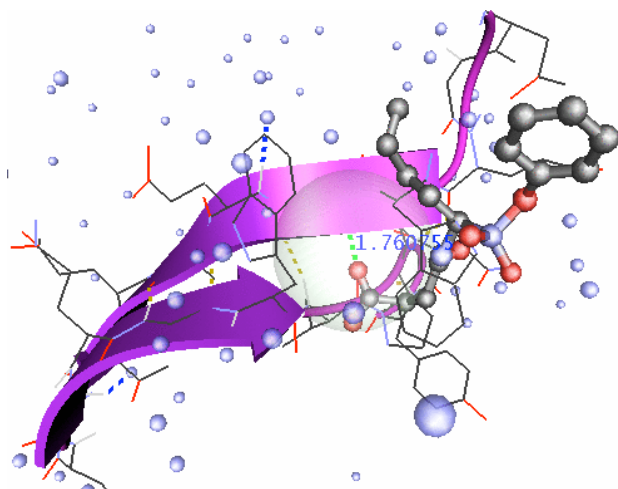


Figure 3. Small blue spheres represent water. Dashed-blue lines represent hydrogen bonds between the ligand and a water molecule. Dashed-green lines represent hydrogen bonds between ligand and receptor. Dashed-yellow lines represent intra-molecular hydrogen bonds. The length of a hydrogen bond is shown along the dashed line. A transparent green sphere highlights the area surrounding a potential hydrogen bond between molecules.

4.1.2 Molecular Overlap Visualization

It is important to consider the intermolecular overlap when scoring the geometric complementarity [17] between entire molecules during the docking process. DockingShop calculates and visualizes atom collisions in real-time during interactive docking to assist users in evaluating the overlap, thus helping to achieve the desired molecular interactions when a protein or ligand is close to a binding pocket. Penetrations are penalized and scored. Nevertheless, our program permits users to ignore atom collisions and rapidly navigate a molecule to a binding site during the interactive docking. A simple grid-based algorithm detects all pairs of atoms inside a protein whose distance is less than 75% of the sum of their van-der-Waals radii. A red sphere of radius proportional to the penetration depth of the two intersecting atoms' van-der-Waals spheres is rendered to visualize the collision. An adapted and expanded grid-based algorithm is used to guide and measure the spatial distance between different molecules. A bounding box based on the geometric shape of structure represents each molecule. The intersection of bounding boxes triggers a real-time collision detection algorithm to

determine the intersection of atoms between involved molecules. A yellow sphere of radius proportional to the penetration depth of the van-der-Waals spheres of the two intersecting atoms shows the collision (Figure 4). These collision detection approaches can also help users reject a solution when the inter- and intra-molecular penetrations exceed a tolerance threshold during visual assessments.

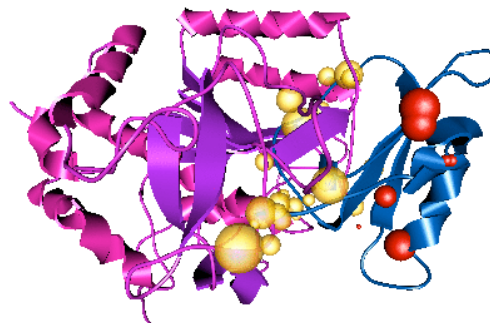


Figure 4. Red spheres represent intra-molecular collisions. Yellow spheres represent inter-molecular collisions. The radius of the collision sphere is proportional to the penetration depth of the van-der-Waals spheres of the two intersecting atoms (PDB code 2SEC).

4.1.3 Energy Calculation and Visualization

Another useful feature is the real-time calculation and visualization of an energy function to help steer the docking process. After an initial location has been determined, energy minimization may be used to find the lowest-energy binding mode of a ligand. It can also contribute significant information to steer the docking process towards identifying the right pocket and binding site. DockingShop provides an interface layer that allows dynamically loading of energy computation modules so that users can couple different energy functions from commercially or internally developed packages. Our program has two schemes for the energy visualization, atom-based and volume-based. In the former, per-atom values are visualized by mapping colors to atoms' van-der-Waals spheres (Figure 5). This approach is useful to rapidly identify high-energy atoms inside the protein. In addition, the program uses a color scheme to visualize individual energy components for a more detailed representation and understanding of the relationship between each energy component and the molecular interaction.

Volume rendering of energy produces a cleaner and more appealing representation than per-atom rendering because it reduces occlusion and visual clutter caused by van-der-Waals spheres. To visualize the energy of a molecule via volume rendering, the program defines a rectilinear voxel block as an axis-aligned bounding volume of the molecule. Then, the program converts the voxels into a 3D texture for hardware-accelerated volume rendering. Scalar energy values associated with atom positions and force field terms are mapped into the voxel block by means of a cubic splatting kernel. Users may define the voxel block in terms of its density in voxels per Angstrom and the radius of a single splat. The splat radius can be uniform or proportional to atom radii. (Figure 5) This feature is further discussed in [24].

DockingShop goes beyond visualization to include numerical analysis of the energy. This feature assists the users to relate calculated quantities to molecular motions and to measure the effect of molecular interactions and structure alignments.

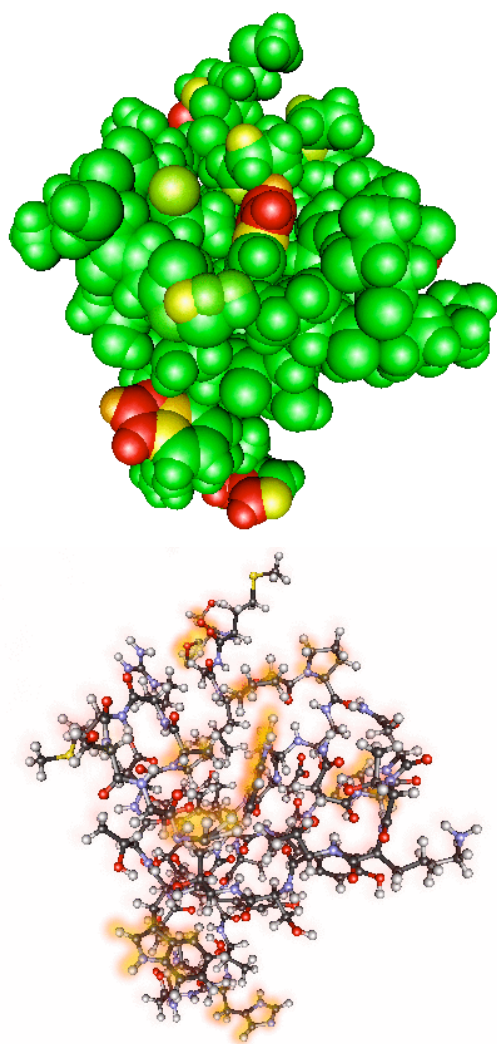


Figure 5. Energy visualization. Top: atom based; bottom: volume based.

4.1.4 Hydrophobic Factor and Other Components

Based on their solubility in water, the residues of the receptor and ligand are classified as hydrophobic or polar. Hydrophobic residues tend to cluster together inside the protein, giving rise to what is known as the hydrophobic effect. Hydrophobicity plays an important role in protein folding as well as in the protein-protein docking process. Three measures of hydrophobicity were implemented in protein-protein docking cases [19] for a “hydrophobicity filter”. We evaluate these measures in our program as part of the adjustable scoring function. The total number of interactions is computed as the sum of three measures including the number of polar-polar, hydrophobic-hydrophobic, and hydrophobic-polar interactions. The hydrophobicity factor is the total of hydrophobic-hydrophobic interactions divided by the total number of interactions. DockingShop also provides a coloring scheme to distinguish hydrophobic and polar residues for the visual assessment to make sure hydrophobic atoms are preferentially buried away from the external environment.

DockingShop provides a labeling system to such display information of protein structures as atom and residue names,

model and chain ID, length of hydrogen bonds and other biochemical properties. These help users to focus on particular regions, motions, interactions, and properties of interest. The dihedral angles of a selected residue may also be retrieved to help users understand and evaluate the protein structure during the manipulation or optimization process.

4.2 Integration of Scoring Function

An “ideal” docking problem solution is to combine the best matching algorithm and scoring schemes [5,17]. The challenge is how to combine them in an efficient manner. An approach used by several docking methods uses coarse-grained scoring to rapidly scan possible solutions and obtain initial “good” candidates, followed by more advanced methods to further discriminate the selected conformations [17]. Recent results have shown that the performance of docking can benefit from the incorporation of biochemical knowledge to steer the process. Therefore, a new research direction integrates the prediction process with biochemical knowledge derived from previous experience [5]. DockingShop integrates biological knowledge into the docking process by allowing users to alter the docking process in response to visualization of computational parameters and scoring functions. This approach helps users to accelerate the search stage by providing them with an efficient method for a reliable discrimination between correct solutions and false positives. The computational parameters described in previous sections can be either visualized or analyzed in numerical form. For example, in addition to the visualization of energy, our program supports on-the-fly calculation of a protein structure's internal energy during manipulation. The energy value is continuously updated during the refinement or manipulation of a protein structure to help users judge the overall quality of the structures and to cull undesirable structures as early in the process as possible.

The program provides a simple adjustable scoring function that is based on a combination of individual parameters with different weights. Users can select which parameters they want to include in the scoring function. The change of weight for each individual parameter is implemented in the form of a slider, which can coherently change the values of the weights either to positive or negative --for a penalty-- in the interactive user interface. The summation of all weighted parameters is the final score. Each parameter and ranking result is continuously updated to aid in evaluating the quality of a docking configuration and to aid in steering the prediction. Users can also couple their own more complex scoring functions into the program through the interface layer, thereby extending the capabilities of DockingShop.

4.3 Interactive Docking of Molecules

An important capability of DockingShop is its six-degree-of-freedom manipulation of single and multiple molecules (like proteins, ligands and water). Users can alter the overall view of a molecular complex through the main window. The orientation and position of each individual molecule inside the complex are controllable by a pilot window. Instead of an exhaustive conformational space search that takes days to complete, users can navigate one molecule to the binding site of another molecule in a very short period of time when binding sites are known. Users can bring two molecules close together and see them bind in real-time. The interaction energy and certain binding properties can be calculated and visualized as described in Section 4.1. From visualization and/or analysis of the energy and binding properties, users can infer the relative strength of the interaction. Usually, the search stage of molecular docking involves different procedures, depending on whether the binding sites are known [17]. When the binding sites are not known, visual guides can help users to

identify a binding site based on their knowledge and intuition. However, in most cases, the binding sites can be predetermined using experimental data and are often found in structural pockets or cavities. DockingShop allows users to use Pocket [25], which is based on an analytical method for detecting pockets in proteins, to decide the binding sites (see Figure 6). When a molecule is moved close to the binding site of another molecule, the orientation of either molecule can easily be changed to sample the conformation space.

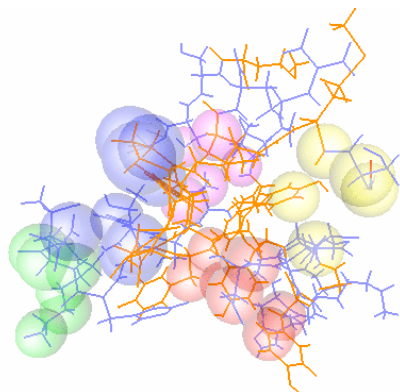


Figure 6. Van-Der-Waals spheres in different colors visualize different pockets.

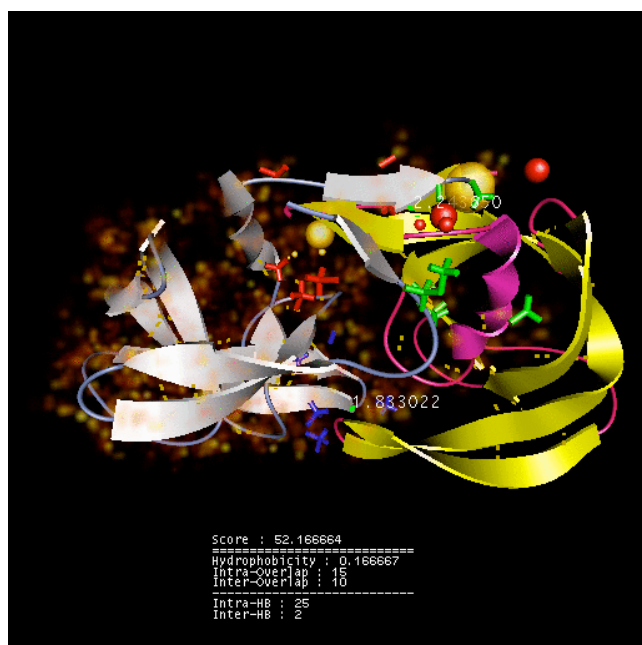


Figure 7. A snapshot of interactive docking. Pockets are visualized by sticks of different colors that represent different binding sites. The value of the score function is calculated. The total score is on the top. The first score is hydrophobicity. The second and third scores are the number of intra- and inter-molecular overlaps. The fourth and fifth scores are the number of intra- and inter-molecular hydrogen bonds (PDB code 4HVP).

The visualization of hydrogen bonds and collision spheres in real time guides users to choose acceptable conformation modes. The ranking of these intermediate modes is based on an adjustable scoring function. Scoring functions may be adjusted based on users' knowledge and used as a comprehensive assessment to filter out many false positive solutions in early stages. When an intermolecular overlap occurs (rendered as a yellow sphere), the affected areas will be highlighted. These overlaps may be resolved by using flexibility of backbone or side chains. In the next section, we discuss how the user can manipulate a molecular structure by moving its backbone, changing its dihedral angles, or replacing its side chains by mutation (see Figure 7 through 10).

4.4 Model Refinement

The purpose of refinement is to improve initial configurations of protein complexes after an estimated binding mode is found. Therefore, an ideal refinement of the receptor structure should allow for flexibility of side chains and backbone dihedral angles. The advantage of having a tool that allows for the flexibility of the backbone is that it allows users to explore the actual transformations that a molecule must undergo in order to properly bind to another molecule. One of the advantages of our program is that it supports interactive real-time manipulation of protein structures without breaking the protein's chemical structure. It provides backbone flexibility for large and small backbone movements allowing conformational changes in the docking process. DockingShop's manipulations are based on those in ProteinShop [7]. They are based on an inverse kinematics (IK) algorithm that transforms parts of a protein with respect to other parts by rotating the backbone dihedral angles, without changing any bond lengths.

To begin manipulation, a user selects a single secondary structure, typically an alpha helix or a beta-strand. The program then renders a *3D manipulation widget* visualized as a translucent green box surrounding the selected structure (see Figure 8). The coil regions connected to the selected secondary structure are activated automatically. The widget may be translated or rotated by dragging it with the mouse. Additionally, a user may activate more coil regions that will serve as buffers for subsequent manipulation because they are highly flexible. In Figure 8, active coil regions are highlighted in yellow.

During refinement, users can also finely adjust the protein structure (receptor) by changing the dihedral angles of a selected residue within the range of the Ramachandran plot [6]. The Ramachandran plot (see Figure 9) presents a bi-varient function of dihedral angles psi and phi that can help users observe the changes in energy resulting from changes in dihedral angle values of a protein during minimization and manipulation. DockingShop shows a 2D scatter plot of the backbone dihedral angles of a selected secondary structure and all active coil regions, overlaid over a pre-computed Ramachandran plot. Dihedral angles of the selected secondary structure are shown as white dots; dihedral angles of all active coil regions are shown as black dots; dihedral angles of the selected residue are shown as a bigger blue dot. Phi angles are mapped on the horizontal axis (from $-\pi$ to π); psi angles are mapped on the vertical axis (from $-\pi$ to π). Dragging the blue dot in the Ramachandran plot window will correspondingly change the dihedral angles of the selected residue (Figure 9).

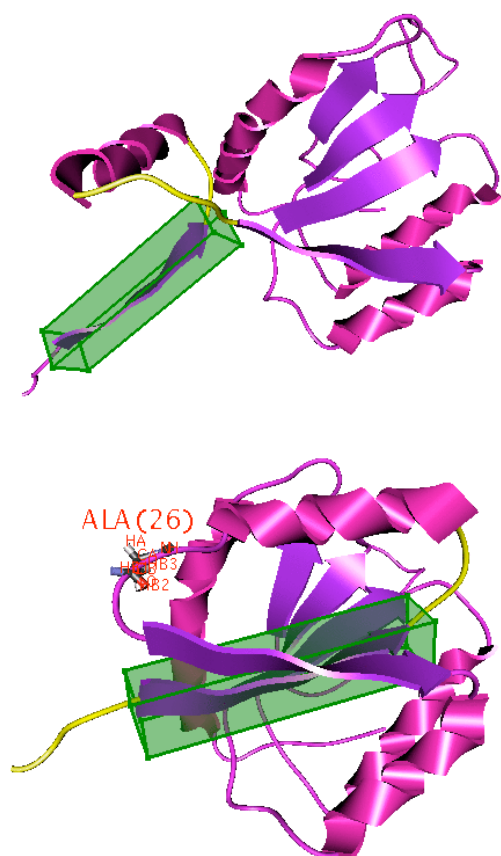


Figure 8. Selected beta strand and surrounding 3D manipulation widget. Selected coil regions are activated for manipulation. Top: before manipulation; bottom: after manipulation. Residue name, ID, and all atom names are labeled along the selected residue.

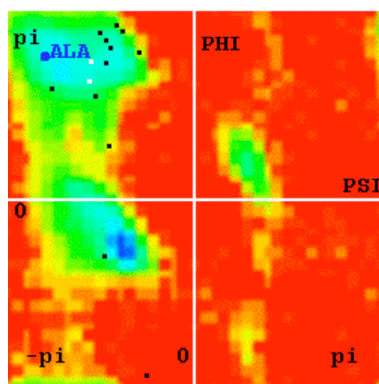


Figure 9. Ramachandran plot. Dihedral angles of the selected secondary structure in Figure 8 are shown as white dots. Dihedral angles of all active regions are shown as black dots. Dihedral angles of the selected residue are shown as a bigger blue dot.

DockingShop permits another type of refinement by substituting the side chain of a selected residue on the receptor protein using a rotamer library [26]. (Dihedral residues on side chains are flexible, and side chain conformations can be classified into a discrete collection of local energy minima called rotamers, listed in a rotamer library.) Certain rotameric states will be higher

in energy than others because of steric interactions that force the side-chain to twist out of the way of neighboring atoms, inflicting a high energy on the residue. The conformation of local backbone remains fixed during the substitution. In addition, as side chain conformations are changed, the hydrogen bonds associated with the side chain are updated accordingly. The flexibility of side chains permits diverse sampling of conformations and may resolve some initially overlaps between receptor and ligand. In addition, DockingShop permits to perform mutations, which allow users to change a residue for another while keeping the backbone fixed.

4.5 Optimization

The native complex is at the global minimum of the free energy change of the formation of the complex relative to its separated components [6]. The determination of the best free energy representation is an open research area. Nevertheless, developing a good approximation to the free energy and finding its global minimum not only provides a powerful approach to solving this problem but also gives additional insight into the thermodynamics involved in macromolecular interactions. The free energy landscape for any protein complex has an enormous number of parameters and local minima; therefore, its global minimization is challenging and requires sophisticated algorithms. Although DockingShop does not provide a global optimization algorithm, it provides an interface to monitor and steer a global optimization algorithm that may be running on a remote machine. In addition, it provides a local optimization function to help create more stable, low-energy conformational complexes that can be used as starting points for a global optimization algorithm. DockingShop allows users to couple different energy functions like AMBER and CHARMM through an interface layer that loads energy computation functions as modules dynamically. Currently, DockingShop provides a module that performs a local minimization of a receptor based on the AMBER force field.

5 CONCLUSION AND FUTURE WORK

We have described an integrated environment --DockingShop-- that helps users to understand and simulate the molecular docking process through a steering mechanism with real-time visual and numerical feedback. It combines molecular structure modeling, interactive docking, and simulation of molecular interactions to create an application for exploration and experimentation in biology. The molecular assemblies created with DockingShop may be used as initial solutions to a more thorough docking algorithm or optimization process. DockingShop's main focus is to provide a graphical interface for integrating human intuition and biological knowledge through the use of interactive molecular docking and adjustable scoring functions. Our work advances computational biology by accelerating the process of molecular docking.

Our current research focuses on the development of interface layers for the integration of user-defined scoring and energy functions as plug-ins. Furthermore, we are investigating the use of stereoscopic rendering and a three-dimensional user interface to reduce the gap between the human's intuition and the interactive molecular docking process. Our future research focuses on the development of more advance features for drug design.

6 ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under contract CHE-0205170 and by the Director, Office of Science, Office of Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

REFERENCE

- [1] Humphrey, W., Dalke, A., and Schulten, K. "VMD – visual molecular dynamics", *Journal of Molecular Graphics* 14, pp.33–38.
- [2] Nelson, M. T., Humphrey, W. F., Gursoy, A., Dalke, A., Laxmikant, V. K., Skeel, R. D., and Schulten, K. "NAMD: A parallel object-oriented molecular dynamics program", *International Journal of Supercomputer Applications and High Performance Computing* 10, 4, pp. 251–268, 1996.
- [3] John Stone, Justin Gullingsrud, Paul Grayson, and Klaus Schulten, "A system for interactive molecular dynamics simulation", In John F. Hughes and Carlo H. Séquin, editors, *ACM Symposium on Interactive 3D Graphics*, New York, pp. 191-194, 2001.
- [4] J. Ding, E. Eskow, T. Lu, W. Liu, L. Jiang, R. Byrd, R. Schnabel, and S. Crivelli, "Protein Structure Prediction Using Physics-based Global Optimization with Knowledge-guided Fragment Packing," *CASP6 – Sixth Meeting on the Critical Assessment of Techniques for Protein Structure Prediction*, pp. A213-A214, 2004.
- [5] Mendez R, Lepae R, De Maria L, Wodak SJ. "Assessment of blind predictions of protein-protein interactions: current status of docking methods". *Proteins*, 52(1): 51-67, Jul 2003.
- [6] Lehninger, A. L., Nelson, D. L., and Cox, M. M. "Principles of Biochemistry", 2nd ed. Worth, New York, New York, 1993.
- [7] Crivelli, S., Kreylos, O., Hamann, B., Max, N. & Bethel, W., "ProteinShop: A tool for interactive protein manipulation and steering", *Journal of Computer-aided Molecular Design*. 18: 271-285, 2004.
- [8] Stefan Birmanns and Willy Wriggers. "Interactive Fitting Augmented by Force-Feedback and Virtual Reality", *J. Struct. Biol.*, 144:123-131, 2003.
- [9] <http://wwwcs.upb.de/~lst/HotDock/index.html>
- [10] <http://www.accelrys.com/insight/>
- [11] Totrov M, Abagyan R., "Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy", *Nat Struct Biol*, 1:259-263, 1994.
- [12] Maxim Totrov, "Protein-protein docking simulation with local backbone flexibility", <http://3dsig.weizmann.ac.il/usersfiles/3dsig/abstracts/7.html>
- [13] Ben-Zeev E, Zarivach R, Shoham M, Yonath A, Eisenstein M, "Prediction of the structure of the complex between the 30S ribosomal subunit and colicin E3 via weighted-geometric docking", *J Biomol Struct Dyn.*, 20(5):669-76, Apr. 2003.
- [14] Roberts, V.A., H.C. Freeman, A.J. Olson, J.A. Tainer, E.D. Getzoff, "Electrostatic orientation of the electron-transfer complex between plastocyanin and cytochrome c", *Journal of Biological Chemistry*, vol. 266, pp. 13431, 1991.
- [15] Sayle, R.A., and Milner-White, E.J. *Trends Biochem. Sci.* 20, pp.374-376, 1995.
- [16] Joël Janin, Kim Henrick, John Moult, Lynn Ten Eyck, Michael J. E. Sternberg, Sandor Vajda, Ilya Vakser, Shoshana J. Wodak, "CAPRI: A Critical Assessment of Predicted Interactions", *Proteins*. Volume 52, Issue 1, pp. 2-9, 2003.
- [17] Inbal Halperin, Buyong Ma, Haim Wolfson, and Ruth Nussinov, "Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions", *Proteins: Structure, Function, and Genetics* 47:409–443, 2002.
- [18] Abagyan, R.A., Totrov, M.M., and Kuznetsov, D.A., "Icm: A New Method For Protein Modeling and Design: Applications To Docking and Structure Prediction From The Distorted Native Conformation", *J. Comp. Chem.* 15, pp.488-506, 1993.
- [19] Norel R, Petrey D, Wolfson H, Nussinov R. "Examination of shape complementarity in docking of unbound proteins", *Proteins*, 35:403–419, 1999.
- [20] <http://pymol.sourceforge.net/>
- [21] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. "UCSF Chimera - A Visualization System for Exploratory Research and Analysis", *J. Comput. Chem.* 25:1605-1612, 2004.
- [22] Anderson A, Weng Z, "VRDD: Applying virtual reality visualization to protein docking and design", *J. Molecular Graphics & Modeling*. 17(3-4): 180-186, 217, 1999.
- [23] Nagata H, Mizushima H, Tanaka H. "Concept and prototype of protein-ligand docking simulator with force feedback technology", *Bioinformatics*, 18(1): 140-6, 2002.
- [24] Clark Crawford, Bernd Hamann, Oliver Kreylos, Silvia Crivelli, "Visualization of force fields in protein structure prediction" Submitted to *IEEE Visualization* 2005.
- [25] <http://www.cs.ucdavis.edu/~koehl/ProShape/>
- [26] S.C. Lovell, J.M. Word, J.S. Richardson and D.C. Richardson, "The Penultimate Rotamer Library", *Proteins: Structure Function and Genetics* 40 pp. 389-408, 2000.

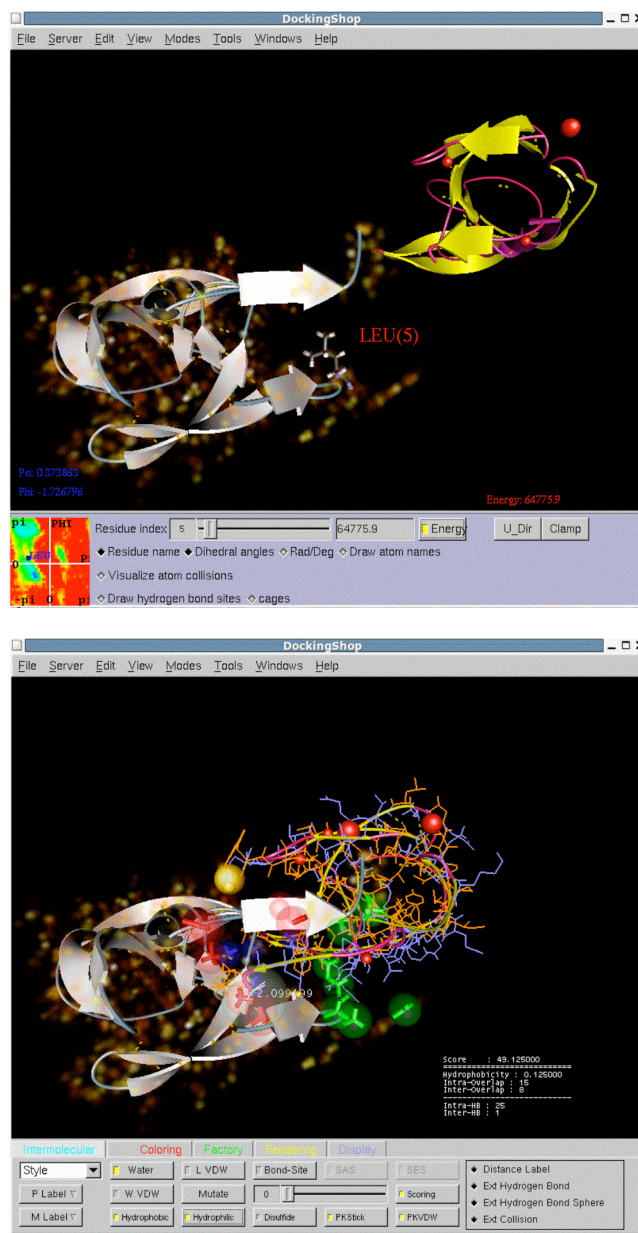


Figure 10. DockingShop's user interface. Top: Manipulation mode; bottom: Visualization mode (PDB code 4HVP).